JOURNAL OF BACTERIOLOGY, Oct. 1995, p. 5393–5400 Vol. 177, No. 19
0021-9193/95/$04.00+0

The Gene for the Longest Known *Escherichia coli* Protein Is a Member of Helicase Superfamily II

NINA BACHER REUVEN,[1] EUGENE V. KOONIN,[2] KENNETH E. RUDD,[2] AND MURRAY P. DEUTSCHER[1]*

Department of Biochemistry, University of Connecticut Health Center, Farmington, Connecticut 06032,[1] *and National Center for Biotechnology Information, National Library of Medicine, Bethesda, Maryland 20894*[2]

Received 15 June 1995/Accepted 2 August 1995

The *Escherichia coli rnt* gene, which encodes the RNA-processing enzyme RNase T, is cotranscribed with a downstream gene. Complete sequencing of this gene indicates that its coding region encompasses 1,538 amino acids, making it the longest known protein in *E. coli*. The gene (tentatively termed *lhr* for long helicase related) contains the seven conserved motifs of the DNA and RNA helicase superfamily II. An ~170-kDa protein is observed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis of ^{35}S-labeled extracts prepared from cells in which *lhr* is under the control of an induced T7 promoter. This protein is absent when *lhr* is interrupted or when no plasmid is present. Downstream of *lhr* is the C-terminal region of a convergent gene with homology to glutaredoxin. Interruptions of chromosomal *lhr* at two different positions within the gene do not affect the growth of *E. coli* at various temperatures in rich or minimal medium, indicating that *lhr* is not essential for usual laboratory growth. *lhr* interruption also has no effect on anaerobic growth. In addition, cells lacking Lhr recover normally from starvation, plate phage normally, and display normal sensitivities to UV irradiation and H_2O_2. Southern analysis showed that no other gene closely related to *lhr* is present on the *E. coli* chromosome. These data expand the known size range of *E. coli* proteins and suggest that very large helicases are present in this organism.

The *rnt* gene encoding RNase T, an enzyme involved in tRNA processing and end turnover (4, 5, 17, 19), is located at approximately 36 min on the *Escherichia coli* genetic map (4). Sequence analysis of the *rnt* gene revealed the existence of a downstream open reading frame (ORF) that is cotranscribed with *rnt* (9). Limited sequence analysis of this ORF (130 amino acids) showed that it contained a high degree of similarity to a common motif found in helicases (8, 23). Because coexpression of an RNA-processing enzyme and a helicase would be unusual and potentially of considerable importance for understanding RNA maturation, we have continued the analysis of this unknown gene.

Helicases are an expanding category of enzymes that participate in many aspects of DNA and RNA metabolism (7). They couple the unwinding of double-stranded polynucleotides to the hydrolysis of ATP. While only about 25 proteins have actually been shown to possess helicase activity, close to 10 times that number have been suggested to be helicases on the basis of analyses of their amino acid sequences (7). This has come about because all the known helicases contain conserved sequence motifs that have predictive value as identifiers for this class of enzymes. The large majority of helicases share seven of these motifs, the details of which have been used to group these proteins into families and superfamilies (7). However, as yet, there is no reliable sequence pattern that can be used to conclusively distinguish between the DNA and RNA helicases.

In this paper, we present the complete sequence of the gene cotranscribed with *rnt*. This gene has been termed *lhr* (for long helicase related). In confirmation of the earlier prediction that this protein might be a helicase (9), we show by sequence analysis that the seven conserved motifs of helicases are indeed present. Remarkably, *lhr* encodes a protein of 1,538 amino acids, making it the longest known protein in *E. coli*. Moreover, the predicted protein of ~170 kDa can be detected in cells overexpressing *lhr*. Despite its size, the Lhr protein is not essential for *E. coli* growth under a variety of conditions, and no phenotype that is associated with null mutations in *lhr* has yet been found.

MATERIALS AND METHODS

Bacterial strains and plasmids. Most of the strains used were *E. coli* K-12 derivatives. Strain UT481 [Δ(*lac-pro*) *hsdS* (r^- m^-) *lacI*^{q} *lacZ*] was used for transformation and plasmid preparations. Strain CF881 (*recB xthA rna*) was used for linear transformation and preparation of DNA for Southern hybridization. *lhr* derivatives of strain CA244 (*trp49 relA1 spoT1 lacZ56 hfr* λ$^+$) were used for growth experiments. The *E. coli* B derivative, strain BL21(DE3) (*hsdS gal ompT*) was used for overexpression of *lhr* under the control of the phage T7 promoter.

Plasmids pBR322, pBR325, and pUC18 were used for cloning. Plasmid pUC4K (Pharmacia) provided the Kan^r cassette used to interrupt *lhr*. Clone λ6F11 from the Kohara library (10), which includes *lhr*, provided DNA for subcloning and DNA sequencing (Fig. 1). The subclones used were the 11-kb *Pvu*II-*Eco*RI fragment cloned into the *Eco*RI-*Eco*RV sites of pBR322 (pBRHL10), the 5.5-kb *Pvu*II-*Pst*I fragment cloned at the *Pst*I and blunted *Eco*RI sites of pBR325, and the 5.7-kb *Bam*HI-*Eco*RI fragment and the 1.4-kb *Bam*HI fragment cloned into pUC18. pBRHL10 was derivatized further by introduction of the 1.26-kb Kan^r cassette from pUC4K into the first and second *Bam*HI sites to generate pBRK1 and pBRK2, respectively.

For expression of the Lhr protein, a fragment beginning 19 nucleotides before the *lhr* coding region (prepared by exonuclease III digestion) and ending at the *Eco*RI site was cloned adjacent to the T7 promoter in plasmid pBS$^+$ (Stratagene). To reduce the loss of the plasmid in BL21(DE3) cells, the insert and T7 promoter were turned around relative to the ColE1 origin of pBS$^+$. The resulting plasmid, termed pBNHEL, was further derivatized by removing the *Bam*HI fragment within *lhr* and replacing it with the Kan^r cassette of pUC4K to generate pBNHELK.

Growth conditions. Cells were grown in either YT or M9 minimal medium (14). For solid media, 2% agar was added. Antibiotics were added at the fol-

* Corresponding author. Present address: Dept. of Biochemistry and Molecular Biology, University of Miami School of Medicine, P.O. Box 016129, Miami, FL 33101. Phone: (305) 243-3150. Fax: (305) 243-3955. Electronic mail address: mdeutsch@medret.med.miami.edu.

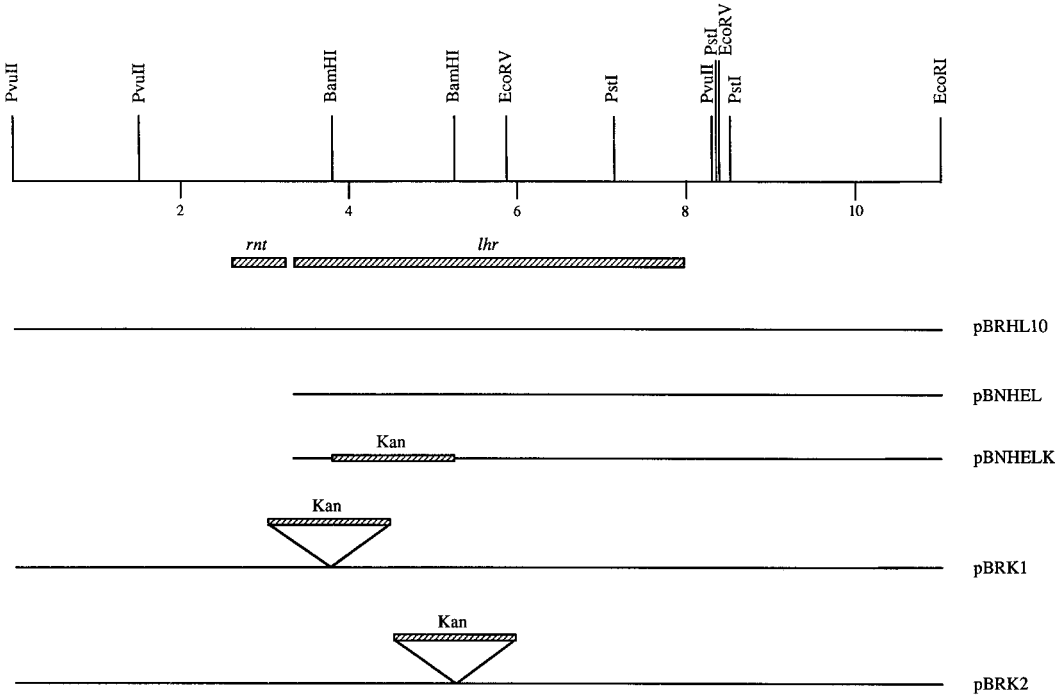

FIG. 1. Restriction map and clones of the *rnt-lhr* region. An 11-kb *Pvu*II-*Eco*RI DNA fragment encompassing the *rnt-lhr* region is shown with the restriction sites used in the course of this work. The various clones used for DNA sequencing, Lhr expression, and physiological studies are also presented. The positions of *rnt* and *lhr*, the Kan^r cassette insertions in pBRK1 and pBRK2, and the deletion insertion in pBNHELK are noted. The *Pvu*II digestion that generated the 11-kb fragment was a partial digestion.

lowing concentrations (in micrograms per milliliter), as needed: ampicillin, 50 to 200; kanamycin, 25; chloramphenicol, 30; and tetracycline, 12.5.

Materials. ^{35}S-dATP and [^{32}P]dATP were obtained from Amersham or Dupont-NEN. Pro-Mix [^{35}S]methionine-cysteine mixture was from Amersham. Sequenase version 2.0 was purchased from U.S. Biochemicals. Restriction enzymes and T4 DNA ligase were obtained from New England Biolabs or American Allied Biochemical. Exonuclease III, nuclease S1, and Pefabloc SC were from Boehringer Mannheim. Klenow fragment DNA polymerase and the DNA probe labeling kit were purchased from Bethesda Research Laboratories. X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) and IPTG (isopropyl-β-D-thiogalactoside) were obtained from Jersey Lab and Glove Supply (Livingston, N.J.). The gel solutions for DNA sequencing were products of National Diagnostics. Gene Screen Plus nylon membranes were from Dupont-NEN. Oligonucleotide primers were synthesized at facilities of the University of Connecticut Health Center or the University of Connecticut. All chemicals used were reagent grade.

DNA techniques. The preparation of plasmid and λ DNAs, cell transformation, cloning, and exonuclease III digestion were carried out according to the standard protocols described by Sambrook et al. (22). Southern blotting and Northern (RNA) blotting were done according to the protocols provided with the Gene Screen Plus membranes. Restriction enzyme digestions were performed according to the manufacturer's instructions. Chromosomal DNA was isolated as previously described (25).

For linear transformation, plasmids pBRK1 and pBRK2 were completely digested with *Eco*RI and *Pvu*II to remove vector sequences. Ten micrograms of linearized DNA was used to transform CF881 cells, selecting for kanamycin-resistant transformants. Transformed cells were restreaked and checked for ampicillin sensitivity to eliminate the possibility of transformation by undigested plasmids. Kan^r Amp^s cells (CF881-K1 and CF881-K2) were used for the isolation of chromosomal DNA for Southern hybridization. The Kan^r interruptions of *lhr* were transferred to strain CA244 by phage P1-mediated transduction.

DNA sequencing. Sequencing was carried out on double-stranded DNA by the dideoxy chain termination method with Sequenase (version 2.0). Forty-two synthetic 18-mer oligonucleotides and the universal and reverse primers were used as primers for completely sequencing both DNA strands. Primers were spaced at 200- to 300-bp intervals along the length of both strands of DNA. The sequence from one primer was used to develop the sequence of the next primer and the primer on the opposite strand. The initial 390 bp of the sequence presented were published earlier (9).

Expression and detection of Lhr protein. A modification of the T7 RNA polymerase procedure of Studier and Moffatt was used (26, 27). BL21(DE3) cells carrying the plasmid plysS (to inhibit T7 RNA polymerase) were transformed with pBS$^+$ based plasmids carrying *lhr* or a Kan^r mutant of *lhr* (pBNHEL and pBNHELK). Transformed cells were maintained on YT-glucose (0.2%) plates supplemented with ampicillin (50 μg/ml), as needed, and chloramphenicol (30 μg/ml). For the induction of *lhr*, fresh colonies were grown in YT-glucose (0.2%) containing high levels of ampicillin (200 μg/ml) and chloramphenicol (30 μg/ml) at 37°C to an A_{550} of 0.2 to 0.3. Cells were collected and resuspended in an equal volume of M9-glucose (0.2%)-ampicillin (200 μg/ml); 2.5 ml of culture was taken for each sample. After growth at 30°C for 30 min, IPTG was added to 1 mM, and growth was continued for another 30 min. Rifampin was then added at 100 μg/ml; an additional 30-min incubation followed. Five microliters (~50 μCi) of ^{35}S-Pro-Mix was then added, and induced cells were incubated for 2 additional h at 30°C. Cells were collected, washed with 2 ml of M9 medium, and resuspended in 0.5 ml of 20 mM Tris-Cl (pH 7.5)–1 mM dithiothreitol–1 mM EDTA–0.2 mM Pefabloc SC–0.5 μg of leupeptin per ml. After sonication for 10 s and centrifugation to remove cell debris, supernatant fractions were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

Computer-assisted sequence analysis. Amino acid and nucleotide sequences were from the SWISS-PROT, PIR, and GenBank databases that are combined in the nonredundant sequence database at the National Center for Biotechnology Information.

Amino acid sequences were compared with those in the nonredundant sequence database by using programs of the BLAST family (1, 2). The BLASTP program was used to screen the amino acid sequence database, and the TBLASTN program was used to screen the conceptual translation of the nucleotide sequence database in six reading frames (1). Compositionally biased regions that tend to produce spurious hits in database searches were excluded from analysis by using the SEG program (1, 29). A modified version of the SEG program was also used to predict nonglobular domains (30). Coiled-coil supersecondary structure was predicted by using the COILS2 program (13). Multiple alignments were constructed by using the MACAW program (24). The statistical properties of amino acid sequences were analyzed by using the SAPS program (3). Protein secondary structure was predicted by using the PHD program (20).

RESULTS AND DISCUSSION

Sequence and analysis of the region downstream of *rnt*. In earlier studies, a 130-amino-acid partial ORF which displayed sequence similarity to some motifs found in helicases was shown to be present downstream of the *rnt* gene (9). We have

completed the sequence of this ORF with a series of primers complementary to previously determined sequences as described in Materials and Methods. Unexpectedly, the ORF was found to be very long, requiring 44 primers to completely determine the sequences of both DNA strands. Approximately 5 kb of DNA sequence, of which 4,617 nucleotide residues represented the ORF, were determined (Fig. 2). In view of the similarity of this ORF to helicases (see below), we have termed this gene *lhr* for long helicase related.

Detailed examination of the nucleotide sequence (Fig. 2) revealed that upstream of the initiator ATG are residues, GGGA, which could serve as a suitable Shine-Dalgarno sequence. Further upstream in the intergenic region between *rnt* and *lhr* are sequences, ATGACA and TATAGT, separated by 18 nucleotides which could be a reasonable σ^{70} promoter (16). Although *rnt* and *lhr* are known to be cotranscribed from a promoter upstream of *rnt*, primer extension analysis indicated that a weak intergenic promoter is also present (9). Interestingly, the 92-nucleotide intergenic region can be folded into two stable stem-loop structures (free energy of −25.6 kcal [1 cal = 4.184 J]) in which both the putative Shine-Dalgarno and promoter sequences are occluded within long stems. Although the physiological significance of this arrangement is not yet understood, the expression of *lhr* may be regulated and an RNA helicase may be necessary to make these sequences accessible.

The derived amino acid sequence of Lhr is also shown in Fig. 2. It contains 1,538 amino acid residues, with a predicted molecular mass of 169,380 Da. The number of amino acids in the coding region would make Lhr the longest protein identified in *E. coli*. A hypothetical protein, YdbA, encoded by a sequence near the replication terminus and interrupted by an insertion element might be longer, but there is no evidence that such a protein is expressed (15). The amino acid composition associated with the derived sequence of Lhr would correspond to an isoelectric point of 6.73. Most amino acids are present in Lhr at levels comparable to those in the average *E. coli* protein. However, the composition of its basic amino acids is quite unusual. The Arg/Lys ratio is close to 6, in contrast to 1 for the average protein. Nevertheless, Arg residues are distributed fairly uniformly over the length of the protein. Other variations from the average protein that were noted are a high tryptophan-to-tyrosine ratio, which would affect the UV spectrum, and a relatively low cysteine content.

Downstream of the *lhr* coding sequence and running in the opposite direction are the C-terminal 73 amino acids of a new *E. coli* gene. Following the provisional nomenclature proposed for uncharacterized *E. coli* ORFs (21), we have named this new gene *ydhD*. Although we have no evidence that YdhD is expressed, sequence comparisons with the protein databases showed statistically significant similarity ($P < 10^{-4}$) to several glutaredoxin-like proteins.

Immediately following the termination codon proposed for the *lhr* gene is a sequence that has the characteristics of a bidirectional *rho*-independent transcription terminator (31). The termination codons of the convergent coding regions of *lhr* and *ydhD* are separated by 50 bp (Fig. 2), and the most prominent RNA secondary structure in this region predicted by MFOLD (32) consists of a palindromic sequence with a 12-bp GC-rich stem and a 6-base loop (free energy of −20.4 kcal). The stem-loop structure is preceded by an A-rich region and is followed by several T residues, as expected for a bidirectional terminator. If this structure functions as a transcriptional terminator, it is in a position to prevent antisense RNAs to the *lhr* and *ydhD* mRNAs from being made. This structure closely resembles the PU* sequence which was previously observed in six intergenic regions separating convergent transcripts and was identified as a bidirectional termination signal (6). The sequence following *lhr* differs from the PU* sequence by just two changes that convert a CG base pair to an AT base pair (residues 4722 and 4743 [Fig. 2]); the base-paired stem is maintained throughout its length. This structure is immediately followed by another sequence that differs from PU* by three changes, but in this case, all of them would disrupt the stem. Further experiments are needed to determine whether the hairpin structure actually functions as a bidirectional terminator.

Relation of the Lhr protein sequence to helicases. Comparisons of the Lhr protein sequence with amino acid sequences in databases showed that its N-terminal region has statistically significant similarity with a number of biochemically characterized and predicted helicases of the DEAD family (8, 23). The highest degree of similarity was with the putative RM62 helicase from *Drosophila melanogaster*, with a probability of matching by chance of less than 10^{-8}.

Figure 3 shows the alignment of the Lhr protein sequence with all putative helicases for which the database search detected *P* values of less than 10^{-4}. Interestingly, all of the sequences listed are from eukaryotes; the similarity between Lhr and the most closely related *E. coli* helicase, DeaD, is below the cutoff level. Lhr contains all seven conserved motifs that are typical of helicase superfamily II (8) without significant deviations from the consensus pattern, except for the conservative substitution of histidine for the distal arginine in motif VI (Fig. 3). Therefore, it is presumed that this protein is a helicase. Despite the fact that Lhr is most similar to the DEAD helicases, it does not contain the signature that gave the name to this family. Instead of DEAD, the sequence DEVH is present in motif II (Fig. 3). A similar observation has been made previously for the *E. coli* RecQ protein, the DNA helicase that is most similar to the DEAD helicases but contains DEAH in motif II. Thus, on the basis of the sequence of this motif alone, Lhr and RecQ would be classified as DEXH proteins (8). The lack of correlation between the motif II signatures and the overall amino acid sequence similarities of these proteins indicates that any classification of helicases based on pattern conservation alone is not adequate (7, 11).

On the other hand, a change in the motif II signature may have important consequences for the actual structure of the helicase domain. A strong correlation between the amino acid signatures in helicase motifs II and VI exists. Proteins that have DEA**D** in motif II contain **H**X_2G$R X_2$R in motif VI; conversely, proteins with DEX**H** in motif II contain **Q**X_2G$R X_2$R in motif VI (7, 8) (amino acid residues for which the correlations have been observed are shown in bold type). Moreover, it has been shown that mutations of the correlated residues in motifs II and VI can partially compensate for each other (18). As can be seen, the correlation between motifs II and VI holds for the Lhr sequence (Fig. 3).

Sequence comparisons alone are insufficient to predict whether Lhr is an RNA helicase or a DNA helicase. Thus, while all the proteins in the DEAD family that have been studied biochemically are RNA helicases (23), RecQ is a DNA helicase (28), despite its closer similarity to the DEAD family than to other DNA helicases. Also, even though the RNA helicases studied in most detail are DEAD proteins (23), there is no significant correlation between the signatures discussed above and the type of nucleic acid that is unwound. For example, the putative RNA helicases of positive-strand RNA viruses contain DEXH/QX_2G$R X_2$R signatures (7, 11).

Of the 1,538 amino acid residues in Lhr, the predicted he-

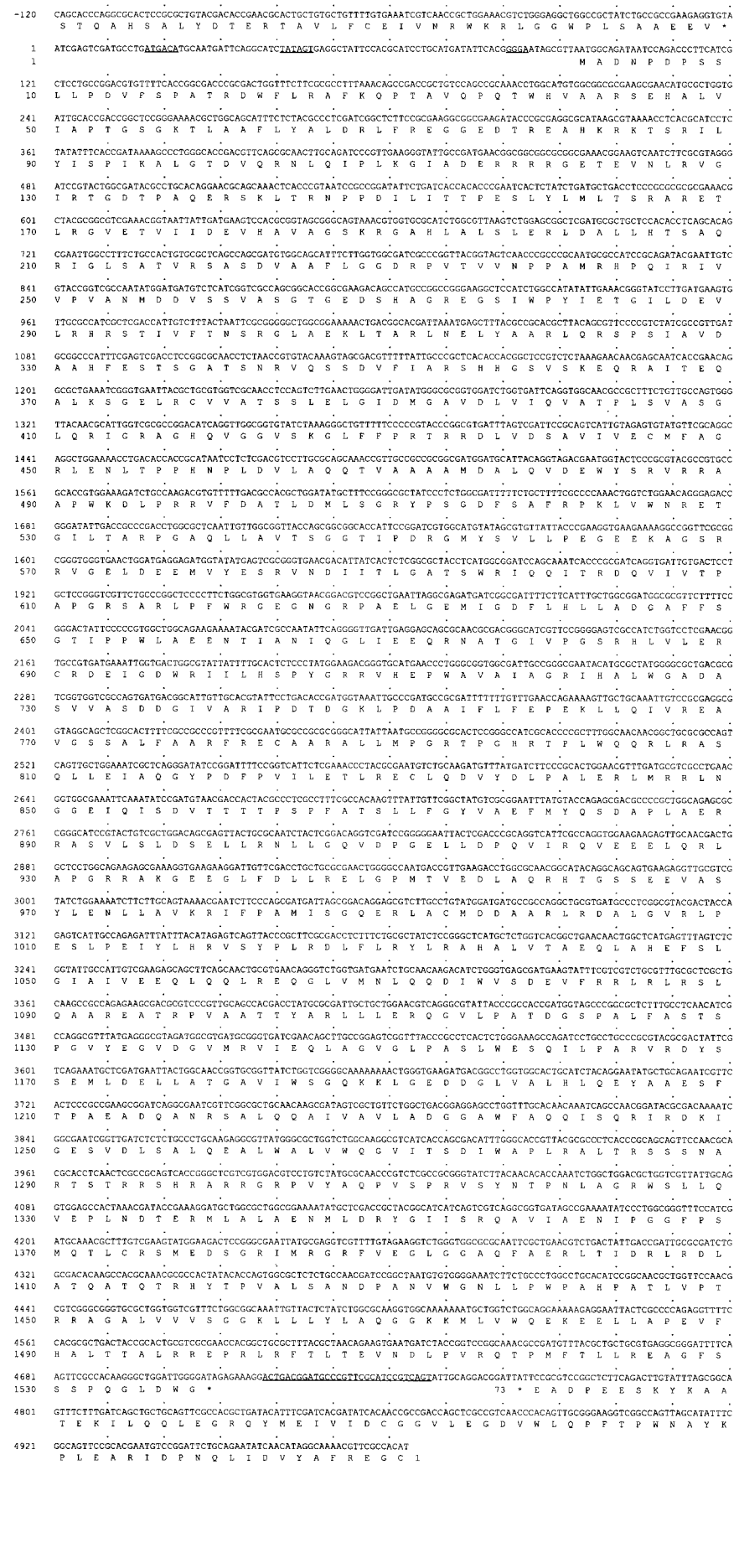

FIG. 2. Sequence of *lhr*. The sequence of ~5 kb of the noncoding DNA strand beginning at the end of the *rnt* coding region and extending through *lhr* to ~0.3 kb past the *lhr* coding region is shown. The numbering of the nucleotide sequence begins at the third nucleotide of the *rnt* termination codon. The derived sequence of Lhr (extending for 1,538 amino acids) is presented in the single-letter format. The putative Shine-Dalgarno sequence and the −10 and −35 regions of a suggested *lhr* promoter are underlined. Also underlined is the putative bidirectional terminator between *lhr* and the convergent glutaredoxin-like gene. The derived amino acid sequence of the C-terminal 73 residues of the glutaredoxin-like gene is also presented. This sequence is derived from the complementary strand.

licase domain consists of about 450 amino acids at the very N terminus. Size variations are generally typical of helicases other than those in the DEAD family, suggesting that many of these proteins contain other domains in addition to the helicase proper (7). The total lengths of the (putative) superfamily II helicases from *E. coli* vary from 403 residues for YejH to 1,538 residues for Lhr (Table 1). Lhr contains by far the longest C-terminal extension.

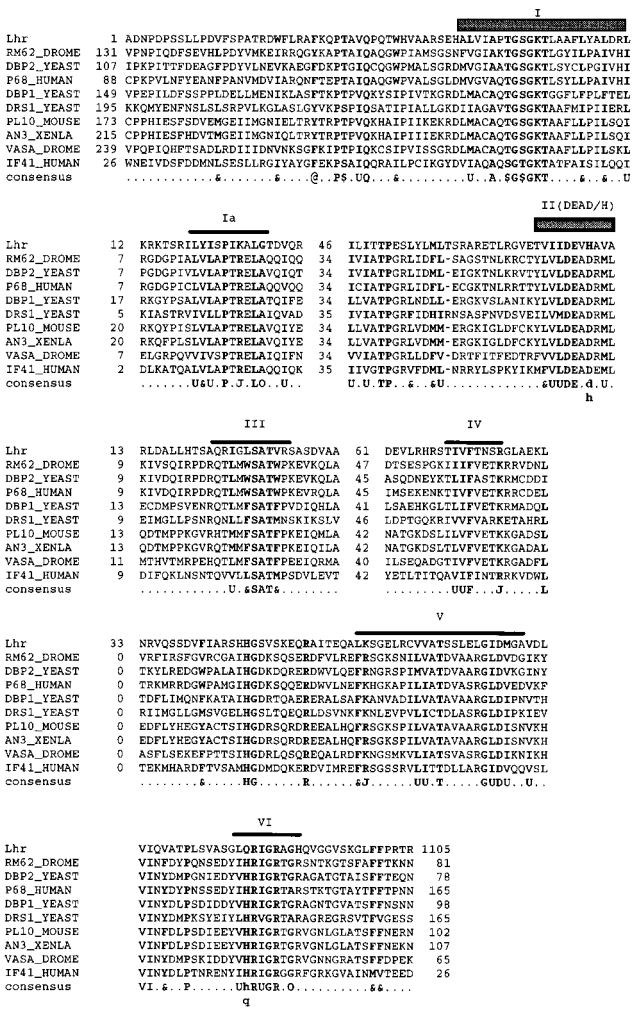

FIG. 3. Conserved motifs in Lhr and related DEAD proteins. The alignment was constructed by using the MACAW program and consists of conserved blocks, with their boundaries determined to achieve maximal statistical significance. The distances between the conserved blocks and from the protein termini are indicated by numbers. The conserved helicase motifs are designated I through VI as described previously (23). Motifs I and II, which are conserved in a vast class of purine NTPases and are thought to directly interact with the NTP substrate (8), are shown by rectangles. The consensus line shows the amino acid residues that are conserved in the 10 aligned sequences. U, a bulky aliphatic residue (I, L, V, or M); @, an aromatic residue (F, Y, or W); &, any of the bulky hydrophobic residues (aliphatic or aromatic); $, a hydroxy residue (S or T); J, a positively charged residue (K or R); O, a small residue (G or A); ⊙, any residue. The DEAD proteins are indicated by their SWISS-PROT identifiers.

The function of the C-terminal part of Lhr, consisting of over 1,100 amino acid residues, remains unknown. However, a comparison with the conceptual translation of nucleotide sequence databases revealed a putative ortholog of Lhr encoded in a *Mycobacterium leprae* genomic cosmid (19a). This presumed coding region, which was not recognized as a gene, is interrupted by several frameshifts, precluding a precise reconstruction; nevertheless, it is clear from the search results that the length of the putative *Mycobacterium* protein is about the same as that of Lhr; moreover, the level of sequence conservation in the C-terminal portion is comparable to that in the N-terminal helicase domain (data not shown). This size and sequence conservation in a phylogenetically distant bacterium suggests that the C-terminal domain(s) of Lhr may have an important physiological function. Database searches with the C-terminal sequences of Lhr and its putative ortholog from *M. leprae* failed to detect significant similarity with any other protein. However, analysis of the Lhr sequence for compositionally biased regions by using the SEG program (29) indicated that this part of the protein, between residues 881 and 1538, probably is nonglobular. A predominantly α-helical secondary structure was predicted for this domain (data not shown). However, the structure of this domain may differ from the classical coiled coil (13), as analysis performed by using the COILS2 program did not predict long coiled-coil regions and database searches failed to show similarities with typical coiled-coil proteins such as myosin.

Southern analysis of *lhr*. By using the entire 4.7-kb *lhr* gene as a probe, Southern analyses of chromosomal DNAs from a variety of strains were carried out to determine whether a gene

TABLE 1. Sizes of domains in putative superfamily II helicases from *E. coli*[a]

Putative helicase	Length (amino acids)			
	Total	N terminus-motif I	Helicase domain	Motif VI-C terminus
Lhr	1,538	46	368	1,124
DbpA	457	42	303	112
DeaD	646	60	295	291
DinG	716	49	621	46
HepA	916	438	306	172
HsdR	1,090	374	232	484
Mfd	1,148	623	282	243
PriA	732	219	379	134
RecG	693	291	296	106
RecQ	610	44	298	268
RhlB	421	48	300	73
RhlE	488	40	309	139
SecA	901	97	491	313
SrmB	444	41	311	92
UvrB	673	34	521	118
YadO	767	22	315	430
YejH	403	26	328	49

[a] The (putative) helicase sequences were selected from the SWISS-PROT database by pattern search as previously described (7).

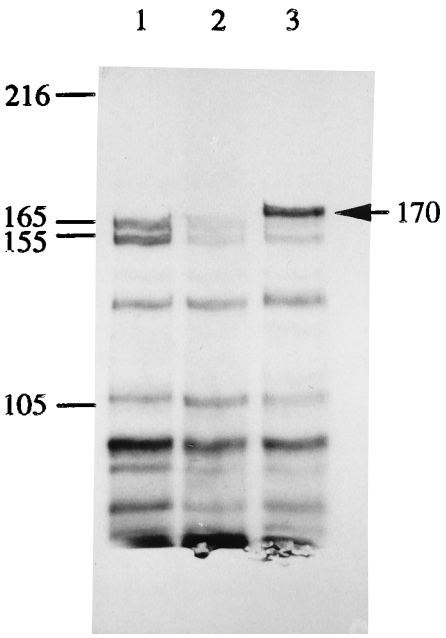

FIG. 4. Expression of Lhr protein. Sonicated extracts, prepared from cells containing no plasmid, pBNHEL, or pBNHELK as described in Materials and Methods, were mixed with sample buffer to give final concentrations of 50 mM Tris-Cl (pH 6.8)–100 mM dithiothreitol–2% SDS–10% glycerol–0.1% bromphenol blue. Samples were boiled for 10 min, cooled on ice for 3 min, centrifuged for 1 min to remove particulate matter, and loaded onto an SDS–5.5% polyacrylamide gel. The gel was run at ~100 V until the dye was ~2 cm from the bottom of the gel. After removal of the dye front and stacking gel, the gel was fixed by soaking in 30% ethanol–10% acetic acid for at least 30 min and then in Amplify (Amersham) for 30 min. After being dried under vacuum at 80°C, gels were exposed to film for ~3 days at −70°C. Lane 1, no plasmid; lane 2, pBNHELK; lane 3, pBNHEL. The positions of size standards (in kilodaltons) are noted. The position of Lhr (170 kDa) is indicated by an arrow.

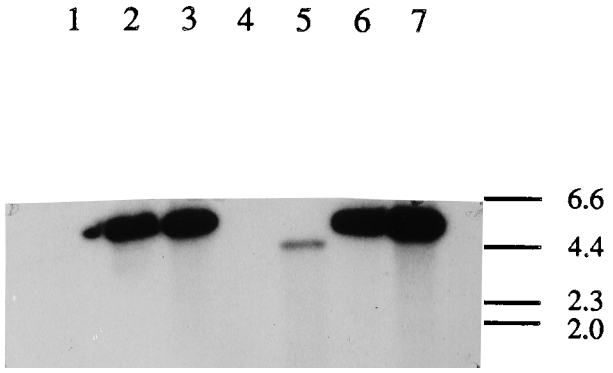

FIG. 5. Southern analyses of chromosomal DNAs from wild-type and kanamycin-resistant cells. Approximately 10 μg each of chromosomal DNA isolated from strain CF881 and its two Kan^r derivatives, CF881-K1 and CF881-K2, and prepared by linear transformation as described in Materials and Methods was completely digested with *Pvu*II and *Eco*RV. The digested DNA was run on a 1% agarose gel with Tris-borate-EDTA buffer containing 0.5 μg of ethidium bromide per ml until the bromphenol blue had migrated 8 cm. Lambda DNA digested with *Hin*dIII was run as a size standard. DNA was transferred to a Gene Screen Plus membrane according to the manufacturer's protocols. The 1.26-kb *Bam*HI fragment from pUC4K was used as the Kan^r probe, and a 0.66-kb *Bam*HI-*Eco*RV fragment from pBRHL10 (Fig. 1) was used as the *lhr* probe. The probes were labeled with [α-^{32}P]dATP by using the random primers DNA labeling system (Bethesda Research Laboratories). Hybridizations carried out according to the Gene Screen protocols were subjected to autoradiography for ~1 h. Lanes 1 and 5, CF881; lanes 2 and 6, CF881-K1; lanes 3 and 7, CF881-K2; lane 4, no sample. Lanes 1 to 3 were hybridized with the Kan^r probe, and lanes 5 to 7 were hybridized with the *lhr* probe. The migration positions of size standards (in kilobases) are shown on the right.

similar to *E. coli* K-12 *lhr* was present. Restriction digests were prepared with *Eco*RI and *Pst*I. Strong hybridization was seen only with DNAs from *E. coli* K-12 and *E. coli* B, whereas weak hybridization was obtained with DNA from *Salmonella typhimurium*. No hybridization was obtained with DNAs from the gram-positive organisms *Bacillus subtilis* and *Micrococcus lysodeikticus* (data not shown). There is also no other gene closely related to *lhr* present on the *E. coli* K-12 chromosome (see Fig. 5).

Expression of Lhr. To establish that *lhr* actually encodes a protein, cells were labeled for 2 h with ^{35}S-Pro-Mix and the resulting radioactive proteins were separated by SDS-PAGE. Although a faint band of 170 kDa was observed, it was not possible to conclusively determine whether this protein was Lhr because interruption of the chromosomal copy of *lhr* (see below) led to only a partial disappearance of this band (data not shown). On the other hand, when cells harboring plasmid pBNHEL, which contains *lhr* under the control of phage T7 RNA polymerase, were labeled with ^{35}S in the presence of rifampin, a prominent band of ~170 kDa was seen; this band was eliminated when *lhr* was interrupted with a Kan^r cassette (Fig. 4; compare lane 3 with lane 2) and when the plasmid was absent (lane 1). These data demonstrate that *lhr* encodes a protein of ~170 kDa and that this protein is relatively poorly expressed, if at all, in cells under the usual growth conditions. Consistent with this conclusion is the observation, based on dot blot analysis, that the *lhr* message is elevated at least 100-fold when pBNHEL is induced (data not shown). Attempts to detect the chromosomal *lhr* message by Northern analysis have been unsuccessful.

Various nucleic acid substrates were examined to determine whether Lhr displayed DNA- or RNA-dependent ATPase activity. Extracts were prepared from cells in which the *lhr* gene on a plasmid was expressed by using the phage T7 RNA polymerase system. The increased expression of Lhr in a parallel culture labeled with ^{35}S was monitored as described above. As controls, cells lacking the plasmid or containing a plasmid with an interrupted *lhr* gene (see below) were used. Despite the large increase in the amount of Lhr protein (Fig. 4), we have not been able to detect an increase in ATPase activity over that of the control with single-stranded, double-stranded, or exonuclease III-treated DNA; poly(U); rRNA or 5S RNA; or a total nucleic acid extract of *E. coli*. Moreover, no differences in ATPase activity between sample and control extracts were observed upon pretreatment with either DNase or RNase. Thus, even though Lhr contains the sequence motifs characteristic of helicases, as yet we have not detected any nucleic acid-dependent ATPase activity associated with this protein. This lack of activity may reflect the use of incorrect substrates or assay conditions or the absence of a required cofactor. Of course, it is also possible that Lhr is not a helicase, but this would be very surprising, considering the close relationship between the presence of helicase motifs in proteins and helicase or nucleic acid-dependent ATPase activity (7).

Interruption of *lhr*. To evaluate the functional importance of *lhr*, we interrupted the chromosomal copy of the gene. A Kan^r cassette was inserted into the plasmid-borne *lhr* gene at either of its two *Bam*HI sites (Fig. 1), as shown by restriction analyses of the resulting clones. Linear transformation and selection for kanamycin resistance were used to transfer each of these interrupted *lhr* genes into the chromosome of strain CF881 to generate CF881-K1 and CF881-K2. Southern blotting of chromosomal DNAs from these two mutant strains and from

CF881 (Fig. 5) indicated that (i) the Kan^r cassette is present in the chromosomes of the mutants but absent from that of strain CF881 (lanes 1 to 3), (ii) the *Pvu*II-*Eco*RV fragment that contains the Kan^r cassette is the same size as the *Pvu*II-*Eco*RV fragment that contains *lhr* (compare lanes 6 and 7 with lanes 2 and 3), and (iii) the size of the *lhr*-containing *Pvu*II-*Eco*RV fragment increases from 4.4 to 5.7 kb upon introduction of the 1.26-kb Kan^r cassette into the chromosome (lanes 5 to 7). These data demonstrate that the chromosomal *lhr* gene has been interrupted by the Kan^r cassette insertions. Moreover, since only one band is observed, the data indicate that no other gene closely related to *lhr* is present on the *E. coli* chromosome.

Analyses of *lhr* mutant strains. On the basis of the recoveries of strain CF881-K1 and CF881-K2 discussed above, it is clear that cells containing an interrupted *lhr* gene remain viable. To examine the physiological consequences of *lhr* mutation in more detail, the gene from strain CF881-K1, which is interrupted in the putative helicase domain, was transferred into strain CA244 by phage P1-mediated transduction. Cells containing the *lhr* mutation grew essentially the same as did wild-type cells on YT plates at temperatures between 22 and 44°C and on M9-glucose plates at 37 or 44°C. In addition, the growth rates of mutant and wild-type cells in YT and YT-glucose media at 37°C were identical (35- and 25-min doubling times, respectively). These data indicate that *lhr* is not required for usual laboratory growth. There was also no difference in the growth of wild-type and mutant strains under anaerobic conditions. Likewise, *lhr* mutant and wild-type cells showed the same rates of recovery from 20- to 23-h incubations in Tris buffer (pH 7.5; starvation for P) and in phosphate-buffered saline (starvation for N and C).

The effect of further stressing the mutant cells by bacteriophage infection was also examined. Phages T3, T4, T7, M13, Qβ, and λ were studied, and all of them plated with equal efficiency on strains carrying mutant or wild-type *lhr* genes. Likewise, the *lhr* mutation had no effect on the efficiency of P1-mediated transduction, and the mutation did not alter the copy number of a pUC plasmid. From these data, it appears that the function of Lhr may be required only under specific physiological conditions not yet tested or that another protein can substitute in its absence.

To test the latter possibility, the *lhr* mutation was combined with mutations in various other genes encoding helicases with the idea that if Lhr is a helicase, then the double mutant might show a growth phenotype. Mutations in the following genes were combined with the *lhr* mutation by phage P1-mediated transduction: *recA*, *recB*, *recD*, *uvrA*, *uvrB*, *uvrD*, *rep*, and *rho*. In each case, the colony size of the double mutant strain was similar to that of the single mutant, indicating that the introduction of Lhr deficiency had no additional effect on cell growth.

We also examined whether strains carrying the *lhr* mutation display altered sensitivities to UV irradiation or H_2O_2. However, we found no increased sensitivity to these treatments under a variety of conditions or in several different genetic backgrounds.

Thus, despite a large number of studies, we have not yet been able to ascertain the function of the *lhr* gene. On the basis of the low-level expression of this gene, it is likely that it functions primarily under specific physiological conditions that we have not reproduced in the laboratory. It would be very surprising, on the basis of sequence analysis of *lhr*, if its gene product was not some type of helicase, given its close relation to this family of enzymes. The size of the *lhr* gene (~0.1% of the *E. coli* genome), and of the protein expressed from it, suggests a significant expenditure of cell resources to maintain and express this gene. It will be of considerable interest to identify the physiological conditions under which *lhr* is expressed and to determine what role its product plays in cell function.

ACKNOWLEDGMENTS

This work was supported by grant GM16317 from the National Institutes of Health to M.P.D.

We thank Shouting Huang for his contributions to the initial phases of this work.

REFERENCES

1. **Altschul, S. F., M. S. Boguski, W. Gish, and J. C. Wootton.** 1994. Issues in searching molecular sequence databases. Nat. Genet. **6:**119–129.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Brendel, V., P. Bucher, I. R. Nourbakhsh, B. E. Blaisdell, and S. Karlin.** 1992. Methods and algorithms for statistical analysis of protein sequences. Proc. Natl. Acad. Sci. USA **89:**2002–2006.
4. **Case, L. M., X. Chen, and M. P. Deutscher.** 1989. Localization of the *Escherichia coli rnt* gene encoding RNase T by using a combination of physical and genetic mapping. J. Bacteriol. **171:**5736–5737.
5. **Deutscher, M. P., and C. W. Marlor.** 1985. Purification and characterization of *E. coli* RNase T. J. Biol. Chem. **260:**7067–7071.
6. **Gilson, E., J. P. Rouset, J. M. Clement, and M. Hofnung.** 1986. A subfamily of *E. coli* palindromic units implicated in transcription termination. Ann. Inst. Pasteur Microbiol. **137B:**259–270.
7. **Gorbalenya, A. E., and E. V. Koonin.** 1993. Helicases: amino acid sequence comparisons and structure-function relationship. Curr. Opin. Struct. Biol. **3:**419–429.
8. **Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov.** 1989. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res. **17:**4713–4730.
9. **Huang, S., and M. P. Deutscher.** 1992. Sequence and transcriptional analysis of the *Escherichia coli rnt* gene encoding RNase T. J. Biol. Chem. **267:**25609–25613.
10. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and storing of a large genomic library. Cell **50:**495–508.
11. **Koonin, E. V.** 1991. Similarities in RNA helicases. Nature (London) **352:**290.
12. **Li, Z., and M. Deutscher.** 1994. The role of individual exoribonucleases in processing at the 3′ end of *Escherichia coli* tRNA precursors. J. Biol. Chem. **269:**6064–6071.
13. **Lupas, A., M. VanDyke, and J. Stork.** 1991. Predicting coiled coils from protein sequences. Science **252:**1162–1164.
14. **Miller, J. H.** 1972. Experiments in molecular genetics, p. 352–355. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
15. **Moszer, I., P. Glaser, and A. Danchin.** 1991. Multiple IS insertion sequences near the replication terminus in *Escherichia coli* K-12. Biochimie (Paris) **73:**1361–1374.
16. **Mulligan, M. E., D. K. Hawley, R. Enriken, and W. R. McClure.** 1984. *Escherichia coli* promoter sequences predict *in vitro* RNA polymerase activity. Nucleic Acids Res. **12:**789–800.
17. **Padmanabha, K. P., and M. P. Deutscher.** 1991. RNase T affects *Escherichia coli* growth and recovery from metabolic stress. J. Bacteriol. **173:**1376–1381.
18. **Pause, A., and N. Sonenberg.** 1992. Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. EMBO J. **11:**2643–2654.
19. **Reuven, N. B., and M. P. Deutscher.** 1993. Multiple exoribonucleases are required for the 3′ processing of *Escherichia coli* tRNA precursors *in vivo*. FASEB J. **7:**143–148.

19a. **Robison, K., and D. Smith.** GenBank accession no. U00012.

20. **Rost, B., and C. Sander.** 1993. Prediction of protein secondary structure at better than 70% accuracy. J. Mol. Biol. **232:**584–599.
21. **Rudd, K. E.** 1993. Maps, genes, sequences, and computers: an *Escherichia coli* case study. ASM News **59:**335–341.
22. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Schmid, S. R., and P. Linder.** 1992. D-E-A-D protein family of putative RNA helicases. Mol. Microbiol. **6:**283–292.
24. **Schuler, G. D., S. F. Altschul, and D. J. Lipman.** 1991. A workbench for multiple alignment construction and analysis. Proteins Struct. Funct. Genet. **9:**180–190.

25. **Silhavey, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions, p. 137–139. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Studier, F. W., and B. A. Moffatt.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. **189:**113–130.
27. **Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff.** 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. **185:**60–89.
28. **Umezu, K., H. Nakayama, and H. Nakayama.** 1990. *Escherichia coli* RecQ protein is a DNA helicase. Proc. Natl. Acad. Sci. USA **87:**5363–5367.
29. **Wootton, J. C.** 1994. Sequences with unusual amino acid compositions. Curr. Opin. Struct. Biol. **4:**413–421.
30. **Wootton, J. C., and S. Federhen.** 1993. Statistics of local complexity in amino acid sequences and sequence databases. Comput. Chem. **17:**149–163.
31. **Wright, J. J., A. Kumar, and R. S. Hayward.** 1992. Hypersymmetry in a transcriptional terminator of *Escherichia coli* confers increased efficiency as well as bidirectionality. EMBO J. **11:**1957–1964.
32. **Zuker, M.** 1994. Prediction of RNA secondary structure by energy minimization. Methods Mol. Biol. **25:**267–294.